

ShapeSonic: Sonifying Fingertip Interactions for Non-Visual Virtual Shape Perception

Jialin Huang
George Mason University
Fairfax, USA
jhuang26@gmu.edu

Alexa Siu
Adobe Research
San Francisco, USA
asiu@adobe.com

Rana Hanocka
University of Chicago
Chicago, USA
ranahanocka@uchicago.edu

Yotam Gingold
George Mason University
Fairfax, USA
ygingold@gmu.edu

ABSTRACT

For sighted users, computer graphics and virtual reality allow them to model and perceive imaginary objects and worlds. However, these approaches are inaccessible to blind and visually impaired (BVI) users, since they primarily rely on visual feedback. To this end, we introduce *ShapeSonic*, a system designed to convey vivid 3D shape perception using purely audio feedback or *sonification*. ShapeSonic tracks users' fingertips in 3D and provides real-time sound feedback (sonification). The shape's geometry and sharp features (edges and corners) are expressed as sounds whose volumes modulate according to fingertip distance. ShapeSonic is based on a mass-produced, commodity hardware platform (Oculus Quest). In a study with 15 sighted and 6 BVI users, we demonstrate the value of ShapeSonic in shape landmark localization and recognition. ShapeSonic users were able to quickly and relatively accurately "touch" points on virtual 3D shapes in the air.

CCS CONCEPTS

• **Computing methodologies** → **Virtual reality; Shape modeling**; • **Hardware** → **Sound-based input / output; Tactile and hand-based interfaces**; • **Human-centered computing** → **Accessibility technologies**.

KEYWORDS

shape, perception, 3D, virtual reality, sonification, non-visual interfaces

ACM Reference Format:

Jialin Huang, Rana Hanocka, Alexa Siu, and Yotam Gingold. 2023. ShapeSonic: Sonifying Fingertip Interactions for Non-Visual Virtual Shape Perception. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23)*, December 12–15, 2023, Sydney, NSW, Australia. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3610548.3618246>

1 INTRODUCTION

3D shape creation and perception are fundamental activities for participation in digital design and virtual worlds. However, interfaces for digital design and exploring virtual worlds are highly visual,

This work is licensed under a Creative Commons Attribution 4.0 International License. Copyrights for third-party components of this work must be honored.



SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0315-7/23/12.
<https://doi.org/10.1145/3610548.3618246>



Figure 1: ShapeSonic allows users to touch virtual shapes in space by sonifying fingertip interactions. The VR headset is used solely for hand tracking and stereo audio. No visuals are shown to the user.

requiring the precise manipulation or perception of 3D shapes on a screen. These approaches are typically inaccessible to people who are blind or visually impaired (BVI) [Mott et al. 2019]. Screen readers and verbal descriptions struggle to convey nuanced 3D shape information. In their place, non-visual 3D shape perception methods are needed.

One direction for incorporating non-visual feedback is through the use of haptics. Touch is most commonly employed for effectively conveying spatial information non-visually. However, technology in tactile displays requires specialized hardware and remains immature and costly [O'Modhrain et al. 2015]. In this work, we investigate *embodied* sonification approaches which can be enabled using commodity hardware, enabling widespread adoption. Sonification methods have been used to effectively convey different types of graphics non-visually, including data visualizations [Siu et al. 2022], 2D shapes [Gerino et al. 2015], and maps [Zhao et al. 2005]. Considerably less work has explored how sound can enhance user's non-visual 3D perception experience. Sighted users may also benefit from our approach as an additional sensory modality for interacting with virtual objects.

We introduce *ShapeSonic*, an interface that enables users to *hear* 3D shapes (Figures 1 and 2). ShapeSonic continuously tracks all the

user's fingertips in real-time, enabling more expressive perception than a single point of contact, or verbal descriptions alone. When the user's fingertips are outside the 3D shape, an ambient sound plays to guide the user's hands to the shape. Fingers contacting the surface play distinct notes on a pentatonic scale, allowing users to distinguish multiple points of contact. To convey geometric details, sharp edges and corners play distinct sounds when touched. The hardware requirements for ShapeSonic are modest, requiring only hand tracking and stereo headphones.

We conducted two experiments to evaluate the effectiveness of ShapeSonic in enabling users to accurately perceive virtual 3D shapes with varying complexity. Our studies involved 6 BVI and 15 sighted users. Our first experiment evaluated the ability of users to recognize shapes. Users were asked to identify shapes from sets of three. Users succeeded in 37 of 45 trials; random guessing would have succeeded in only 15 trials. Our second experiment evaluated users' ability to precisely locate various keypoints or landmarks on a shape (e.g. the ear, feet, and tail of a cat) through interaction. Users were, on average, 6× more accurate using ShapeSonic than from a verbal description of the shape.

In summary, our contributions are:

- ShapeSonic, an embodied sonification approach for non-visual 3D shape perception that runs on off-the-shelf commodity hardware.
- A novel technique for *hearing* shapes using a combination of tones in ambient space and surface contact sounds which consider the underlying geometric features.
- Results demonstrating the effectiveness of ShapeSonic in supporting shape recognition and understanding of spatial relationships. We conducted user studies with $N = 6 + 15$ BVI and non-BVI participants.

2 RELATED WORK

Previous computational approaches to non-visual representations of graphics have explored tactile and auditory media. Tactile approaches [Bau et al. 2010; Benko et al. 2016; Fang et al. 2020; Giudice et al. 2012; Jansson et al. 2003; Peters 2011; Sinclair et al. 2019; Xu et al. 2011; Yem et al. 2016] are promising, but require specialized hardware—most of it not commercially available. Pen-based haptic devices typically only present a single point of contact rather than allowing whole-hand interaction. Performance in haptic recognition tasks is significantly improved with more than one point of contact [Jansson et al. 2003].

Haptic shape displays are another alternative to displaying 2D and 2.5D media [Bornschein et al. 2015; Siu et al. 2019] and support whole-hand interaction, but these devices still require very specialized hardware. Other tactile approaches for conveying graphical information involve the creation of a physical 2D or 3D shape [Furferi et al. 2014; Fusco and Morash 2015; Karbowski 2020; Li et al. 2011; Panotopoulou et al. 2020; Shi et al. 2019, 2017; Stangl et al. 2015]. While tactile interactions with shapes are effective [Klatzky et al. 1985], these methods involve a lengthy fabrication process and are less suitable for real-time feedback and interaction. Our proposed approach allows users to use their entire hands to explore a 3D shape that can be dynamically rendered.

Auditory approaches have used computer vision to trigger spoken descriptions in response to user pointing [Fusco and Morash

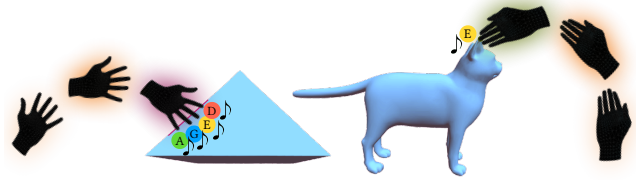


Figure 2: Illustration of ShapeSonic. Users' hands are guided to shape surfaces with sonified distance. Fingertips contacting the shape play notes in a pentatonic scale. Sharp edges and corners are sonified with distinct sounds. These shapes were used in our landmark localization experiment (Section 4). Users explored the shapes with ShapeSonic, and then were asked to localize features.

2015] or considered automatic sonification of datasets. Gerino et al. [2015] explored several sonification techniques to map 2D bitmap images to the amplitude and frequency of sine waves, evaluating the ability of users to discriminate between several simple 2D shapes (triangle, square, diamond, circle). Other related approaches have also considered or converted 2D data as vector graphics, such as lines and areas [Su et al. 2010; Yoshida et al. 2011]. Users were able to successfully comprehend and reproduce a variety of shapes after less than an hour of training. Many approaches have sonified 1D data such as time series [Brewster et al. 2002; Holloway et al. 2022; Sharif et al. 2022; Siu et al. 2022; Zhao 2006]. Alonso-Arevalo et al. [2012] applied 1D sonification techniques to the cross sections of 3D shapes; they measured success as a 1D data perception task. These approaches are largely focused on sonifying 1D and 2D data and cannot be directly generalized to communicate 3D shapes, the goal of ShapeSonic.

Heed et al. [2015] computed the echolocation of 3D surfaces and reported that users enjoyed the experience, but not whether users were able to comprehend the surface. Echolocation requires extensive training, and it is not clear the extent to which echolocation can be used to comprehend shapes beyond localization, distance, density, and discriminating based on size, texture, and contour [Andrade et al. 2018, 2021; Milne et al. 2014; Norman et al. 2021; Olmos and Cooperstock 2012; Thaler and Goodale 2016; Wallmeier and Wiegrebe 2014]. Expert echolocators have been shown to be approximately 75% accurate at distinguishing between a shape whose 2D contours are square, equilateral triangle, horizontally oriented rectangle, or vertically oriented rectangle [Milne et al. 2014]. ShapeSonic aims to convey accurate perception of more complex 3D shapes non-visually.

3 METHOD

The ultimate display would, of course, be a room within which the computer can control the existence of matter.

Ivan Sutherland [1965]

The goal of ShapeSonic is to convey a vivid sense of a 3D shape using purely audio feedback or sonification. For the purposes of this research, we deliberately eschewed verbal interactions, such as labeling shapes and announcing what someone touched. ShapeSonic

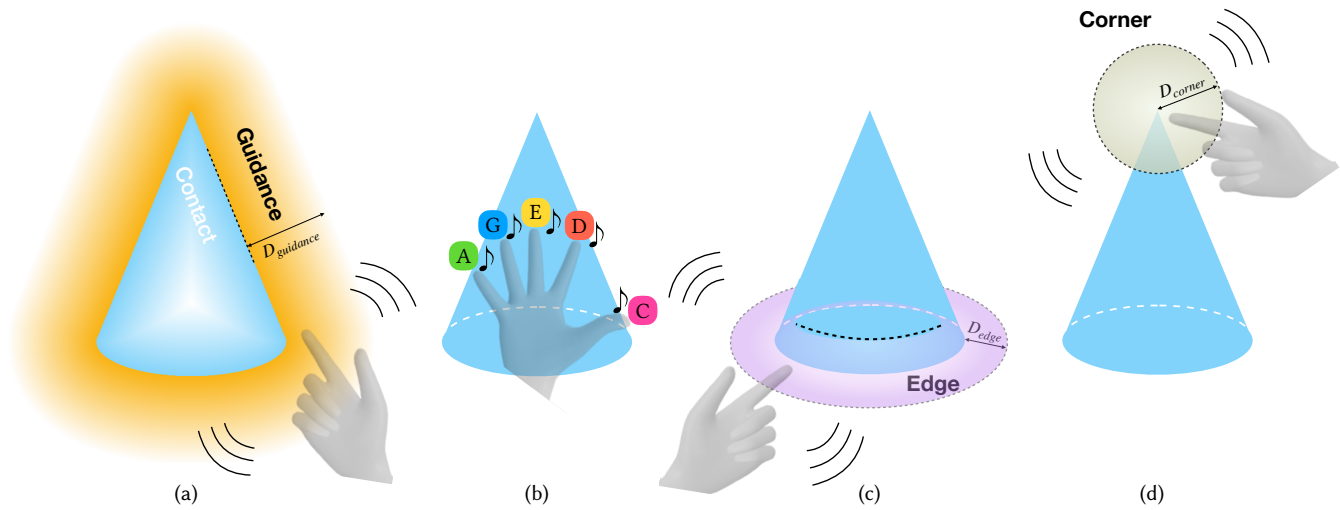


Figure 3: An overview of user actions and the resulting sonifications. (a) ShapeSonic sonifies the space around a shape to guide the user's hands to contact it. The Guidance sound gets louder as the hands approach the shape. ShapeSonic sonifies the shape's interior with a Contact sound, loudest at the surface and then decreasing in volume to the interior. (b) Each fingertip plays the Contact sound at different pitches of a pentatonic scale. ShapeSonic also sonifies fingertips near a sharp edge (c) or corner (d).

is designed to convey percepts that can't easily be conveyed verbally. Our design goals were to substitute (1) the tactile sensations of hand-shape contact and (2) visual information conveying the shape's position at a distance. We arrived at our proposed design by studying the literature on non-visual shape perception, exploratory prototyping, early prototype review with a BVI individual, and feedback from a pilot study (Section 4).

We based our design on tracking the user's hands with respect to a virtual 3D object and reacting with sound. Hand tracking is an active area of research, with commercially available and affordable implementations. Stereo audio output is affordable and universally available. The central design question is *how to map from hands in space to sound in ears*. Sonification approaches in the literature have considered many sound attributes for mapping continuous data, such as pitch, loudness, tempo, attack, and modulation. Among these, pitch and loudness are the most accurately perceived [Sharif et al. 2022; Walker 2002, 2007; Walker and Kramer 2005; Wang et al. 2022]. We focus our continuous sound design on these two attributes. In terms of sound localization, humans perceive direction quite accurately ($< 10^\circ$), but distance quite inaccurately (positively correlated under ideal circumstances) [Middlebrooks 2015]. However, even directional accuracy would be insufficient to distinguish a user's two hands in close proximity—as they could be when touching the same shape—let alone the fingertips of the same hand. As a result, we did not explore 3D positional sound sources. Instead, to ensure an unambiguous mapping from the user's hands, we sonify the left and right hand in the left and right audio channel, respectively. In addition, we play all left-ear sounds at a lower pitch than right-ear sounds.

See Figure 3 and the supplemental video for an overview of user actions and sonifications. Our approach assumes solid shapes with a well-defined interior and exterior. In our experiments, we focus on a single 3D object, although our approach is general.

Sonification regions. In ShapeSonic, space is primarily divided into *outside* the shape and *inside* the shape. (There are two smaller layered regions near sharp edges and corners.) So long as the user's hands are within half a meter of the shape, the user will constantly experience a *guidance* sound. The shape's surface divides space into two disjoint regions. To mirror the binary sensory transition from not touching to touching [Dellon et al. 1992], a different *contact* sound plays, with an abrupt transition, when any fingertip is inside the shape. The shape's sharp edges and corners also play distinct sounds if any fingertip is close enough. If, for example, the user is touching the corner of a cube, they will hear three sounds: the contacting, edge, and corner sound.

We utilized a professional sound designer to create intuitive and pleasant sounds. The guidance sound is ambient. The contact sound is contrasting. The abrupt change from the guidance to contact sound was designed to mimic the abrupt physical sensation of fingertips contacting a surface. The edge sounds resemble a plucked guitar string. Corner sounds resemble a bell. The sounds are all easily recognized as distinct.

Distance is loudness. As one of the two most accurately perceived sonification properties, we mapped distance from the surface (edge, corner, resp.) to loudness. Sounds are loudest when touching or almost touching the surface (contact and guidance), edge, or corner. Sounds get quieter as fingertips move further away. (For the contact sound, further away means deeper inside the shape.) For example, the guidance sound gets louder as the user's hand gets closer to the shape, in effect guiding the user's hand to the shape. For the guidance, edge, and corner sounds, the closest fingertip determines the volume. For the contact sound, each fingertip in contact activates a sound independently (see below). Our design is as if fingertips have microphones and sound sources are located on the surface of



Figure 4: Tutorial Shapes. When familiarizing themselves with ShapeSonic, formal study participants interacted with a torus, a non-convex smooth shape, and a triangular prism, a flat-sided shape with sharp edges and corners.

the shape (one facing inward and one facing outward), along edges, and at corners.

We use an exponential falloff for the volume of each sound. This function was chosen and refined through preliminary testing with both BVI and sighted people. It is steep enough to convey the direction to the surface, edges, and corners. The precise formula is $V_{\text{region}}(x) = a + be^{-\frac{5x}{D_{\text{region}}}}$, where x is the distance, a is the minimum volume, b controls the steepness, and D_{region} is the largest distance for which the sound plays. Table 1 in the supplemental materials contains our parameters for the four sound types.

Fingertip contact maps to a pentatonic scale. Each fingertip in contact with the shape produces a distinct musical note. Pitch and loudness are the most accurately perceived sonification properties. We reserved pitch for fingertips, since the same sound cannot be layered with different loudnesses. This gives the user the ability to detect whenever a fingertip comes in and out of contact with the surface of the shape. A user can explore space with all five fingertips and, by rotating their hand and listening for musical notes, understand the surface orientation. Alternatively, the user can explore with one fingertip outstretched and the rest folded; when the outstretched fingertip makes contact, the user can unfold the other fingertips to discover the surface orientation.

We generate the distinct notes as different pitches of the contact sound. We chose the pitches of a pentatonic scale, which is a five-note scale often used in various musical traditions worldwide. Notably, when multiple notes from the pentatonic scale are played together, they are commonly perceived as harmonious rather than dissonant. We used a major pentatonic scale (e.g., CDEGA) corresponding to the user’s thumb, index, middle, ring, and little fingers, respectively. Even though most users won’t perceive an *absolute pitch*-to-finger mapping [Deutsch 2013], most people accurately perceive relative pitch [Wier et al. 1977], allowing them to determine which fingers are newly in contact with the surface. Multi-finger contact allows users to feel the local curvature properties of the shape directly.

3.1 Implementation

Implementing ShapeSonic requires reliable hand tracking, stereo headphones, and fairly light computing needs. We chose the widely available and affordable Oculus Quest 1 VR headset, which has

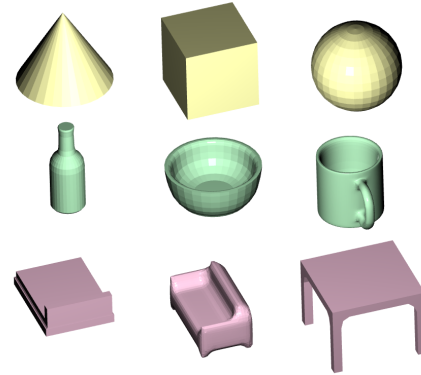


Figure 5: Shape Recognition Task. In our formal study, participants were given a description of three shapes in a set (corresponding to rows above), and then interacted with one of the three at random. Afterwards, participants were asked to identify the shape.

hand tracking, stereo sound, and a sufficiently powerful processor. We covered the screens inside the headset with a physical block, preventing users from seeing the screens. We experimented with several alternative hand tracking approaches. We found MediaPipe [Zhang et al. 2020] (an off-the-shelf RGB-based library) to be inaccurate. We found the Leap Motion controller’s tracking region too small. We found the Sensoryx gloves not sufficiently stable to compensate for its lack of wide availability. In the future, we expect our hardware requirements to decrease further. RGB-based hand tracking is an active research area and may soon be possible on any computer or phone.

We implemented ShapeSonic’s interface in Unity using C#. The software loads a shape stored as a signed distance field (SDF) along with a set of 3D sharp edges and corners. We chose to use SDFs as our shape representation, since it allows us to directly access the distance to an object’s surface at the cost of increased memory usage. The alternative would be to compute point-to-mesh distances on-the-fly, which can be slow for concave or complex meshes, particularly on the Oculus Quest 1. Thus, we opt to compute the SDF for a mesh in an offline pre-processing step. We used the Python SDF¹ library for computing an SDF from a mesh and libigl [Jacobson et al. 2018]’s `sharp_edges` function for detecting sharp edges and corners. The threshold was tuned per model so output edges matched common sense. For cases where the automatic edge and corner detection failed, we manually labeled edges and corners using Blender. Automatic edge and corner detection can fail for meshes with too sparse or dense sampling. In sparse regions, dihedral angles may be large due to low sampling. In dense regions, a seemingly sharp edge may be slightly beveled and never have large dihedral angles.

4 EVALUATION

We conducted studies with 15 sighted and 6 BVI users to evaluate the effectiveness of ShapeSonic on 3D shape perception

¹SDF 0.3.5 <https://pypi.org/project/SDF/>

tasks. We were unsure as to the limits of ShapeSonic, so we took two precautions. First, we designed tasks with progressive difficulty. Even if ShapeSonic users failed at the challenging tasks, they might still have succeeded at the easier ones. This would give us insight as to ShapeSonic’s limits. Second, prior to recruiting from the BVI population, which is limited, we ran a *pilot study* with 6 sighted users, who are more easily recruited. This allowed us to refine ShapeSonic and our experimental protocol before conducting a larger, *formal study* with 9 sighted and 6 BVI users. Per the requirements of our IRB, participants in both studies completed the user study in-person at our university campus. Participants were compensated with a \$25 Amazon gift card and reimbursed for any transportation fees.

4.1 Overview of Tasks

We designed two types of tasks to evaluate shape perception. These stayed largely the same between our pilot and formal studies. The first task was *shape recognition*, where users were asked to identify shapes in sets of threes (Figure 5). We expected identifying a shape to be possible even without a clear or precise 3D shape percept. One of the sets was composed of simple geometric primitives, which we expected most users to succeed in. Accuracy and recognition time were recorded as performance metrics. All shapes were positioned in the same 3D location and uniformly scaled to lie within a 40cm bounding box. The precise protocol was revised for the formal study to make the results easier to analyze. The precise set of objects was refined as well based on participant confusion.

The second task was a *landmark localization task*. Users were given a verbal description of an object, including its dimension and orientation in 3D space. As a baseline, users were first asked, without sonification, to guess landmark locations (e.g., the ear of a cat) in 3D space based purely on the verbal description. We then enabled sonification. After a period of free exploration, users were asked to locate the same landmarks again. The verbal baseline was not part of the pilot study. It was added to the formal study to provide a numerical measure of improvement. We expected this to be a challenging task evaluating 3D shape comprehension and recall. We used a simple shape (pyramid), for which we expected success, and a complex shape (dog or cat) to test the limits of ShapeSonic. For the pyramid, the landmarks were the top corner, bottom corner (any of the four), and anywhere on the bottom face.

Before beginning the tasks, participants were given time to familiarize themselves with ShapeSonic. After completing all tasks, participants answered a short questionnaire to collect feedback.

4.2 Pilot Study

We recruited 6 sighted participants for a pilot study to assess the effectiveness of ShapeSonic and guide the study design for our formal evaluation (Section 4.3). Participants comprised 3 women and 3 men and were 27.5 years old on average. Participants always began with a tutorial in which we described basic strategies as they interacted with a sphere and a torus for 3 minutes each. (The torus is visible in Figure 4.) We chose the sphere for its simplicity and the torus because it is smooth but non-convex.

4.2.1 Shape recognition. For the shape recognition task, participants were given 2 sets of objects, each with 3 shapes (bowl, mug,

and bottle; table, bed, and sofa). These can be seen as the bottom two rows of Figure 5. Participants had up to 5 minutes of free exploration time and were then asked to recall the order in which the 3 shapes were shown. Results are shown in Table 5. In total, participants ordered 12 sets of three shapes (6 participants \times 2 sets of shapes). Among these 12 trials, participants correctly identified all three shapes 5 times, one shape 6 times, and zero shapes 1 time. In contrast, 12 random guesses would have resulted in a distribution of 2, 6, and 4, respectively.² A χ^2 test suggests that our participants did significantly better than random ($p = 0.034$). Based on this evidence of success, we adjusted the experimental protocol for our formal study’s shape recognition task to be (a) more difficult and (b) easier to analyze. Instead of interacting with and ordering all shapes in a set, users would interact with and identify one randomly chosen (and counterbalanced). We also adjusted some of our shapes as a result of user feedback. Users were surprised by some of the shapes’ corners and edges (e.g. concave corners and parallel sharp edges along the lip of the mug). These were algorithmically detected (Section 3.1). For our revised study, we created a smoother mug and manually removed concave corners from all shapes.

4.2.2 Landmark localization. The landmark localization task was performed on a pyramid and a dog. Participants were given 5 minutes of free exploration time with a shape, and then asked to locate landmarks (up to 1 minute for each). For the dog, the landmarks were its nose, feet (any of the four), and any part of the tail. Participants were given a verbal description of each shape’s dimensions. Results are shown in Table 2. Overall, users were able to roughly perform landmark localization on the two shapes. For our formal study, to separate the effect of verbal descriptions from our sonification, we revised our study protocol to record and compare the position accuracy with and without sonification.

4.2.3 ShapeSonic Refinements. There were some differences in the version of ShapeSonic used for the pilot study. First, the contact sound did not vary in pitch between fingers. Second, only the guidance sound’s volume modulated according to distance. The remaining sounds (contact, corner, edge) played at a constant volume. Our final design was informed by feedback from pilot study participants.

We also observed that some users needed more time to learn the sonification and effective exploration strategies. In our formal study protocol, we extended the training period and replaced the sphere with a triangular prism.

4.3 Formal Study

For our formal experiment, we recruited a total of 15 participants. Sighted participants comprised 7 women and 2 men and were 25 years old on average. BVI participants comprised 4 women and 2 men and were 55.5 years old on average. Among the 6 BVI participants, 3 are totally blind, 2 have minimal remaining vision in one eye, and one has vision in one eye that cannot be corrected to normal. The study protocol began with a 10-minute tutorial and

²When ordering sets of three, there is exactly one correct answer, 0 answers with two objects in the correct position, 3 answers with one correct position, and 2 answers with zero correct positions. Since 50% of random guesses have two mistakes, any partial success is virtually impossible to distinguish from chance.

training period in which we guided participants to hear all forms of sonification and taught them basic strategies:

- Use both hands to explore the boundaries of the shape along the three egocentric axes.
- Use a single fingertip along a straight line to detect surface extents. During this process, avoid contact with the other fingers. Their contact sounds can be identified by their differing (pentatonic) pitches.
- After reaching one point of contact with the shape, try to follow the surface to understand its shape (e.g., round or flat). Multiple fingertip contact is helpful.

During the training period, participants interacted with a triangular prism and a solid torus. We chose the prism because it makes use of all sonifications in our system (including corners and edges) yet has simple, flat faces.

4.3.1 Shape recognition. For our revised shape recognition task, we created three sets of shapes (Figure 5): { cube, sphere, cone }, { bowl, mug, bottle }, { table, bed, sofa }. Each set was a trial in which we briefly described the three shapes, selected one at random, and gave participants 5 minutes to interact with it. Participants were then asked to identify the shape. Every shape appeared five times in the experiment, 3 times by sighted and 2 times by BVI participants.

Trial results can be found in Table 3. The 9 sighted participants correctly identified 22 shapes out of their 27 trials. Random guessing would have resulted in only $\frac{1}{3}$ correct answers (9 in this case). A χ^2 test shows that this result is extremely unlikely to arise randomly ($p = 10^{-7}$). The 6 BVI participants correctly identified 15 shapes out of their 18 trials. A χ^2 test shows that this is also extremely unlikely due to chance ($p = 10^{-5}$). The success rate for sighted and BVI participants was similar (81% versus 83%). We conclude from the data that ShapeSonic is effective in allowing both sighted and BVI users to identify shapes.

4.3.2 Landmark localization. We revised our landmark localization trials to include a paired control. We also believe that the large, real-world variability in dog shapes was a source of confusion. As cats have a more uniform shape than dogs, we replaced the dog with a cat in our revised landmark localization study. After a detailed verbal description of a shape (pyramid and then cat), we asked participants to touch each landmark with sonification disabled. This established a baseline for comparison. We then enabled sonification, gave participants 5 minutes to interact, and then asked them to relocate each landmark in under 1 minute, respectively. The short relocation time was designed to activate proprioceptive recall. (We did not strictly enforce the 1 minute timer. Some participants reported trouble keeping their hands in the air for a long time because of age and body limitations. We allowed them to rest.)

Table 4 and Figure 6 show the improvement in landmark positioning error when using sonification. Sonification significantly improved landmark positioning accuracy for all features. The average overall improvement was 9.6 cm or a factor of 6.2 \times . 81 of 90 trials showed improvement. We found no significant overall differences between sighted and BVI participants. Two BVI participants, P11 and P15, showed exceptional improvement when using our sonification. Both were BVI from birth.

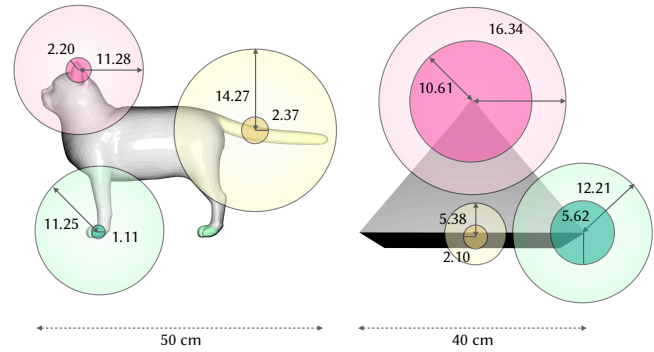


Figure 6: Average localization error with with and without sonification for the cat and pyramid. Units are centimeters. Landmarks: The ear, feet, and tail (any part) of the cat; the top corner, any bottom corner, and bottom face of the pyramid. Landmarks are shown in magenta, green, and yellow, respectively. The smaller, darker circles depict the resulting error after exploration.

4.4 Questionnaire

At the conclusion of our pilot and formal study, participants answered a short questionnaire to collect feedback. We asked two structured Likert-scale questions, “Please place the experience on a continuum between someone describing a shape verbally and feeling a shape physically,” and “Please rate the degree to which you perceived the sensation of 3D shapes.” The results can be seen in Figure 7. On average, participants placed ShapeSonic almost exactly between “Verbal shape description” (1) and “Feeling a physical shape” (5) with an average of 2.9 ± 0.5 . On average, participants also reported perceiving the sensation of 3D shapes approximately halfway between “not at all” (1) and “very vividly” (5) with an average of 3.6 ± 0.6 . Despite the participant variability, we consider these to be encouraging results given that ShapeSonic allows participants to interact with and obtain a spatial understanding of virtual shapes *with absolutely no visual or tactile feedback*.

4.5 Comparisons

We are not aware of a comparable 3D sonification system to ShapeSonic. Still, we can compare ShapeSonic with other tactile and sonic shape recognition experiments.

Klatzky et al. [1985] provides a best-case comparison for our shape recognition task. Participants handled physical objects, and so had whole-hand haptic feedback. They reported accuracy of 96% with almost all (94%) responses given in <5 seconds. 100 objects were used. ShapeSonic users were 82% accurate, with a median response time of ~2 minutes. Objects were known to participants in sets of 3.

Gerino et al. [2015] evaluated 2D sonification techniques and found 75% and 77% (sighted and BVI) accuracy when distinguishing between sets of four simple polygons. Alonso-Arevalo et al. [2012] sonified properties of slices of a 3D object as 1D functions. This was, in effect, an experiment for sonifying the shape of 1D functions. They evaluated recognizing 1D functions or identifying minima/maxima/inflection points along the curve.

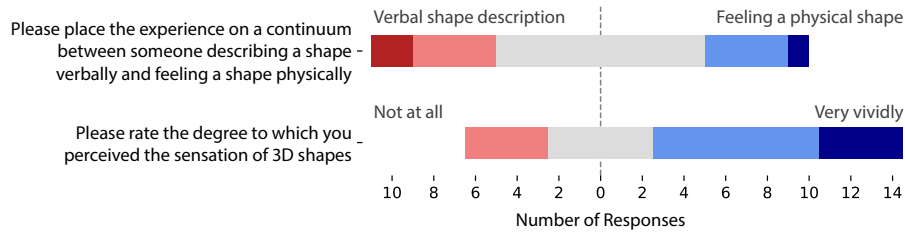


Figure 7: A diverging stacked bar chart showing participants in our pilot and formal studies placed ShapeSonic on a continuum between verbal descriptions and physically interacting with a shape.

It is unclear how to extend 2D sonification approaches to 3D shape recognition and landmark localization. How would the user pick a 3D slice (plane orientation and offset)? Many slice silhouettes are rather uninformative, as can be seen when viewing CT or MRI data. (For example, a slice through a chest would produce an oval shape.) A landmark point in 3D will almost surely never lie on a 2D slice. What distance should be sonified, the distance to the 2D silhouette or 3D distance to the shape? These are interesting questions, and could lead to alternative approaches to 3D sonification.

The plane selection difficulty is related to the projection direction in 2D tactile graphics. Panotopoulou et al. [2020] performed a shape recognition experiment among a set of 5 distinct objects and found a large variation in success depending on viewpoint (from 6% to 21%). A second shape recognition task involved sets of 3 shapes with subtle differences; their improved approach achieved 58% accuracy versus 29%.

4.6 Observations, Discussion, and Future Work

Although some participants reported using ShapeSonic to be similar to feeling a physical shape and with a very vivid 3D shape sensation (Figure 7), we do not believe that participants would have succeeded without at least being given the (verbal) name of the object they were interacting with. Participants relied on both the brief verbal descriptions and ShapeSonic to accurately sense the shapes. We posit that when using ShapeSonic, participants imagine the shapes in their minds and then verify and refine their mental model with embodied interaction. It appears to have been relatively easy for participants to verify if the shape they envisioned matched the one sonified by ShapeSonic. At times, however, we observed a mismatch between the geometry that participants envisioned based on a given shape description and the actual geometry that was sonified. For instance, some participants imagined a standing cat with only its hind legs on the ground, whereas the sonified shape was of a cat standing with all four paws on the ground. This mismatch resulted in some participants taking more time to perceive the correct geometry or failing to understand the shape they were exploring. These observations are in-line with prior work on sonification strategies which show that a small description followed by a systematic exploration strategy helps users better contextualize the information and support their accurate interpretation [Brewster et al. 2002; Siu et al. 2022; Zhao 2006].

Overall, the sound design used by ShapeSonic was perceived as intuitive and satisfying by most participants but required some learning and familiarization. Some participants initially struggled

to remember all the sounds associated with each region, and therefore, they needed to be reminded of the meaning of each sound. Some participants requested more sound feedback to provide additional information, while others struggled to distinguish the four sounds used in the ShapeSonic. Over the course of the study and especially during the later landmark localization tasks, most participants were able to connect the sounds with their relative meanings. This suggests that we have not reached the training ceiling for users of ShapeSonic. Similar to other kinds of graphics, with sonification, good strategies need to be learned and training can take time [Hermann 2002; Zhao et al. 2005].

During our user study, one participant (P3) displayed exceptional proficiency in utilizing the sound feedback provided by ShapeSonic to sense the shapes accurately. P3 utilized all her fingertips and capitalized on the volume change design for corner and edge sounds. Notably, she explored the shapes with her own strategy, moving in a slow and steady pace, and did not miss any of the shapes presented by ShapeSonic. These findings suggest that certain users may possess unique abilities to effectively leverage the sound feedback provided by ShapeSonic, which may inform future design considerations for such technology, such as codifying their successful strategies or targeting such users with designs that a general audience would find too complex.

In our user study, three participants reported that their hands grew tired and that the VR helmet felt heavy. In the future, we can ensure that virtual objects are sonified atop a table, so that users can rest their forearms or elbow as they explore. External hand-sensing hardware, such as a laptop webcam, would allow users to use ShapeSonic with only headphones instead of a heavy VR setup.

The hand tracking latency forces users to move slower than they otherwise might. With future, lower latency hand tracking, we would be able to evaluate its effect. Extremely low latency interactions may provide a sensation akin to “ear haptics,” where the tactile sensation is transferred.

We would also like to explore sonification of additional surface attributes, such as curvature, material (e.g., wood, glass, fur), softness [Lau et al. 2018], geometric texture (rough versus smooth) [Tymms et al. 2018], and even color. We envision users being able to enable or disable multiple sonification channels. Future explorations could also incorporate additional dynamic verbal descriptions in reaction to hand interactions as well. These descriptions could complement the feedback provided through sonification. We would also like to explore the effect of ShapeSonic as an additional sensory channel to visual perception. In the future we are interested in extending

ShapeSonic to perceive and convey *motion* and other dynamic effects such as deformations that change over time.

In addition to richer sonification design for a single object, we would like to explore sonifications of more complex objects, including arrangements of multiple objects in a scene. While our method is general, in this work, we only presented users with one object at a time. There might be additional sound interactions that need to be provided to enable accurate perception of multiple objects and their relationship.

Lastly, we would like to expand ShapeSonic to a digital shape fabrication system that supports a feedback loop of perception-editing. The initial shape could be retrieved from a public 3D shape dataset, or generated using a pretrained large language model via verbal description. The user could then feel the shape using ShapeSonic and could make incremental edits to refine the shape. This would extend our system to not only support 3D shape perception, but also creation.

5 CONCLUSION

We introduced ShapeSonic, a system for BVI users to hear virtual 3D shapes using their hands. ShapeSonic provides users with embodied sonifications of their fingertip interactions with 3D shapes. We designed sounds which are intuitive and distinguishable, like a guitar sound for edges and a bell for corners. In addition, we used properties of sounds (pitch and volume) to convey rich information to users. We evaluated the effectiveness of ShapeSonic through a user study with 15 sighted and 6 BVI users. Our evaluation consisted of an object recognition task and a landmark positioning task. Both sighted and BVI users were able to perceive 3D shapes of varying complexity and, on average, found the experience halfway between a verbal description and feeling a shape physically. We designed ShapeSonic to rely on commodity hardware. Our hope is that our method lowers access barriers for BVI users to participate in 3D shape perception and design activities.

ACKNOWLEDGMENTS

We are grateful to our anonymous reviewers and many user testers, formal and informal, throughout the process. Gene Kim provided important early feedback to let us know we were on the right track. Henro Kriel suggested using pitches on a pentatonic scale for distinguishing per-finger contact. Kim Avila helped us immensely with outreach for our user study. Authors Huang and Gingold were supported by a gift from Adobe Inc. Author Hanocka was supported by gifts from Adobe Inc, Google, and the United States National Science Foundation (IIS-2304481 and CNS-2241303).

REFERENCES

- Miguel A. Alonso-Arevalo, Simon Shelley, Dik Hermes, Jacqueline Hollowood, Michael Pettitt, Sarah Sharples, and Armin Kohlrausch. 2012. Curve shape and curvature perception through interactive sonification. *ACM Transactions on Applied Perception* 9, 4 (Oct. 2012), 17:1–17:19. <https://doi.org/10.1145/2355598.2355600>
- Ronny Andrade, Steven Baker, Jenny Waycott, and Frank Vetere. 2018. Echo-house: exploring a virtual environment by using echolocation. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. ACM, Melbourne Australia, 278–289. <https://doi.org/10.1145/3292147.3292163>
- Ronny Andrade, Jenny Waycott, Steven Baker, and Frank Vetere. 2021. Echolocation as a Means for People with Visual Impairment (PVI) to Acquire Spatial Knowledge of Virtual Space. *ACM Transactions on Accessible Computing* 14, 1 (March 2021), 1–25. <https://doi.org/10.1145/3448273>
- Olivier Bau, Ivan Poupyrev, Ali Israr, and Chris Harrison. 2010. TeslaTouch: electrovibration for touch surfaces. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*. 283–292.
- Hrvoje Benko, Christian Holz, Mike Sinclair, and Eyal Ofek. 2016. Normaltouch and textretouch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In *Proceedings of the 29th annual symposium on user interface software and technology*. 717–728.
- Jens Bornschein, Denise Prescher, and Gerhard Weber. 2015. Collaborative creation of digital tactile graphics. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 117–126.
- S Brewster, L Brown, R Ramloll, and W Yu. 2002. Browsing Modes For Exploring Sonified Line Graphs. In *Proc. of the 16th British HCI Conference London*, Vol. 2. 2–5.
- Evan S. Dellon, Robin Mourey, and A. Lee Dellon. 1992. Human Pressure Perception Values for Constant and Moving One- and Two-Point Discrimination. *Plastic and Reconstructive Surgery* 90, 1 (July 1992), 112. https://journals.lww.com/plasreconsurg/Citation/1992/07000/Human_Pressure_Perception_Values_for_Constant_and.17.aspx
- Diana Deutsch. 2013. Absolute pitch. In *The Psychology of Music (Third Edition)*. Elsevier Academic Press, San Diego, CA, US, 141–182. <https://doi.org/10.1016/B978-0-12-381460-9.000005-5>
- Cathy Fang, Yang Zhang, Matthew Dworman, and Chris Harrison. 2020. Wireality: Enabling Complex Tangible Geometries in Virtual Reality with Worn Multi-String Haptics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–10. <https://doi.org/10.1145/3313831.3376470>
- Rocco Furferi, Lapo Governi, Yary Volpe, Luca Puggelli, Niccolò Vanni, and Monica Carfagni. 2014. From 2D to 2.5 D ie from painting to tactile model. *Graphical Models* 76, 6 (2014), 706–723.
- Giovanni Fusco and Valerie S Morash. 2015. The tactile graphics helper: providing audio clarification for tactile graphics using machine vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 97–106.
- Andrea Gerino, Lorenzo Picinali, Cristian Bernareggi, Nicolò Alabastro, and Sergio Mascetti. 2015. Towards large scale evaluation of novel sonification techniques for non visual shape exploration. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 13–21.
- Nicholas A. Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M. Kramer. 2012. Learning non-visual graphical information using a touch-based vibro-audio interface. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*. ACM Press, Boulder, Colorado, USA, 103. <https://doi.org/10.1145/2384916.2384935>
- Tobias Heed, Johanna Möller, and Brigitte Röder. 2015. Movement induces the use of external spatial coordinates for tactile localization in congenitally blind humans. *Multisensory research* 28, 1-2 (2015), 173–194.
- Thomas Hermann. 2002. *Sonification for exploratory data analysis*. Ph. D. Dissertation.
- Leona M Holloway, Cagatay Goncu, Alon Israr, Matthew Butler, and Kim Marriott. 2022. Infsonics: Accessible Infographics for People who are Blind using Sonification and Voice. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–13. <https://doi.org/10.1145/3491102.3517465>
- Alec Jacobson, Daniele Panozzo, et al. 2018. libigl: A simple C++ geometry processing library. <https://libigl.github.io/>.
- Gunnar Jansson, Massimo Bergamasco, and Antonio Frisoli. 2003. A new option for the visually impaired to experience 3D art at museums: manual exploration of virtual copies. *Visual Impairment Research* 5, 1 (2003), 1–12.
- Caroline Karbowski. 2020. See3D: 3D Printing for People Who Are Blind. *Journal of Science Education for Students with Disabilities* 23, 1 (Feb. 2020). <https://doi.org/10.14448/jsesd.12.0006>
- Roberta L. Klatzky, Susan J. Lederman, and Victoria A. Metzger. 1985. Identifying objects by touch: An “expert system”. *Perception & Psychophysics* 37, 4 (July 1985), 299–302. <https://doi.org/10.3758/BF03211351>
- Manfred Lau, Kapil Dev, Julie Dorsey, and Holly Rushmeier. 2018. A Human-Perceived Softness Measure of Virtual 3D Objects. *ACM Transactions on Applied Perception* 0, 0 (2018).
- Nan Li, Zheshen Wang, Jesus Yuriar, and Baoxin Li. 2011. Tactileface: A system for enabling access to face photos by visually-impaired people. In *Proceedings of the 16th international conference on Intelligent user interfaces*. 445–446.
- John C. Middlebrooks. 2015. Sound localization. In *Handbook of Clinical Neurology*. Vol. 129. Elsevier, 99–116. <https://doi.org/10.1016/B978-0-444-62630-1.00006-8>
- Jennifer L Milne, Melvyn A Goodale, and Lore Thaler. 2014. The role of head movements in the discrimination of 2-D shape by blind echolocation experts. *Attention, Perception, & Psychophysics* 76 (2014), 1828–1837.
- Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stoakley, and Meredith Ringel Morris. 2019. Accessible by design: An opportunity for virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 451–454.
- Liam J. Norman, Caitlin Dodsworth, Denise Forestiere, and Lore Thaler. 2021. Human click-based echolocation: Effects of blindness and age, and real-life implications in a 10-week training program. *PLOS ONE* 16, 6 (June 2021), e0252330. <https://doi.org/10.1371/journal.pone.0252330> Publisher: Public Library of Science.

- Adriana Olmos and Jeremy R Cooperstock. 2012. MAKING SCULPTURES AUDIBLE THROUGH PARTICIPATORY SOUND DESIGN. (2012).
- Sile O'Modhrain, Nicholas A Giudice, John A Gardner, and Gordon E Legge. 2015. Designing media for visually-impaired users of refreshable touch displays: Possibilities and pitfalls. *IEEE transactions on haptics* 8, 3 (2015), 248–257.
- Athina Panotopoulou, Xiaoting Zhang, Tammy Qiu, Xing-Dong Yang, and Emily Whiting. 2020. Tactile line drawings for improved shape understanding in blind and visually impaired users. *ACM Transactions on Graphics* 39, 4 (Aug. 2020). <https://doi.org/10.1145/3386569.3392388>
- Benjamin J Peters. 2011. *Design and fabrication of a digitally reconfigurable surface*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- Ather Sharif, Olivia H. Wang, and Alida T. Muongchan. 2022. "What Makes Sonification User-Friendly?" Exploring Usability and User-Friendliness of Sonified Responses. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Athens Greece, 1–5. <https://doi.org/10.1145/3517428.3550360>
- Lei Shi, Holly Lawson, Zhuohao Zhang, and Shiri Azenkot. 2019. Designing Interactive 3D Printed Models with Teachers of the Visually Impaired. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland UK, 1–14. <https://doi.org/10.1145/3290605.3300427>
- Lei Shi, Yuhang Zhao, and Shiri Azenkot. 2017. Designing Interactions for 3D Printed Models with Blind People. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 200–209. <https://doi.org/10.1145/3132525.3132549>
- Mike Sinclair, Eyal Ofek, Mar Gonzalez-Franco, and Christian Holz. 2019. Capstan-Crunch: A Haptic VR Controller with User-supplied Force Feedback. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 815–829. <https://doi.org/10.1145/3332165.3347891>
- Alexa Siu, Gene SH Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting Accessible Data Visualization Through Audio Data Narratives. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
- Alexa F. Siu, Son Kim, Joshua A. Miele, and Sean Follmer. 2019. shapeCAD: An Accessible 3D Modelling Workflow for the Blind and Visually-Impaired Via 2.5D Shape Displays. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. Association for Computing Machinery, Pittsburgh, PA, USA, 342–354. <https://doi.org/10.1145/3308561.3353782>
- Abigale Stangl, Chia-Lo Hsu, and Tom Yeh. 2015. Transcribing across the senses: community efforts to create 3D printable accessible tactile pictures for young children with visual impairments. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 127–137.
- Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal de Lara, and Khai N Truong. 2010. Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 17–26.
- Ivan E Sutherland. 1965. The Ultimate Display. In *Proceedings of the IFIP Congress*, Vol. 2. New York, 506–508.
- Lore Thaler and Melvyn A. Goodale. 2016. Echolocation in humans: an overview. *WIREs Cognitive Science* 7, 6 (2016), 382–393. <https://doi.org/10.1002/wcs.1408> [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1408](https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1408)
- Chelsea Tymms, Esther P. Gardner, and Denis Zorin. 2018. A Quantitative Perceptual Model for Tactile Roughness. *ACM Transactions on Graphics* 37, 5 (Oct. 2018), 1–14. <https://doi.org/10.1145/3186267>
- Bruce N. Walker. 2002. Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied* 8, 4 (2002), 211–221. <https://doi.org/10.1037/1076-898X.8.4.211>
- Bruce N. Walker. 2007. Consistency of magnitude estimations with conceptual data dimensions used for sonification. *Applied Cognitive Psychology* 21, 5 (2007), 579–599. <https://doi.org/10.1002/acp.1291> [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.1291](https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.1291)
- Bruce N. Walker and Gregory Kramer. 2005. Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception* 2, 4 (Oct. 2005), 407–412. <https://doi.org/10.1145/1101530.1101534>
- Ludwig Wallmeier and Lutz Wiegrebe. 2014. Self-motion facilitates echo-acoustic orientation in humans. *Royal Society Open Science* 1, 3 (2014), 140185.
- R. Wang, C. Jung, and Y. Kim. 2022. Seeing Through Sounds: Mapping Auditory Dimensions to Data and Charts for People with Visual Impairments. *Computer Graphics Forum* 41, 3 (June 2022), 71–83. <https://doi.org/10.1111/cgf.14523>
- Craig C. Wier, Walt Jesteadt, and David M. Green. 1977. Frequency discrimination as a function of frequency and sensation level. *The Journal of the Acoustical Society of America* 61, 1 (Jan. 1977), 178–184. <https://doi.org/10.1121/1.381251>
- Cheng Xu, Ali Israr, Ivan Poupyrev, Olivier Bau, and Chris Harrison. 2011. Tactile display for the visually impaired using TeslaTouch. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. 317–322.
- Vibol Yem, Ryuta Okazaki, and Hiroyuki Kajimoto. 2016. FinGAR: combination of electrical and mechanical stimulation for high-fidelity tactile presentation. In *ACM SIGGRAPH 2016 Emerging Technologies*. 1–2.
- Tsubasa Yoshida, Kris M Kitani, Hideki Koike, Serge Belongie, and Kevin Schlei. 2011. EdgeSonic: image feature sonification for the visually impaired. In *Proceedings of the 2nd Augmented Human International Conference*. 1–4.
- Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. <https://doi.org/10.48550/arXiv.2006.10214> arXiv:2006.10214 [cs].
- Haixia Zhao. 2006. *Interactive sonification of abstract data: framework, design space, evaluation, and user tool*. University of Maryland, College Park.
- Haixia Zhao, Catherine Plaisant, and Ben Shneiderman. 2005. iSonic: interactive sonification for non-visual data exploration. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*. 194–195.

A PILOT STUDY PROTOCOL

I. Review the Informed Consent Form.

- i. Review the Health and Safety Warnings for Oculus Quest VR device. Individuals at risk per the Health and Safety Warnings should not participate.
- ii. Obtain verbal consent.

II. Explain: "We have created a program for touching virtual 3D shapes sonification or sound feedback to let you know if your hands are near or shape. Our goal is to evaluate whether our program can allow you to distinguish 3D shapes and how accurately you can perceive the shape. This study will take approximately 1 hour. In the study, you will use our program to touch various objects."

III. Direct the participant to wear the VR headset and run the program.

IV. Procedure:

- i. Training (6 minutes). Have the user interact with a sphere and then a torus. Tell them they will have up to 3 minutes to interact with each shape.
 1. Explain the kinds of sonification: "Our project provides four types of sound feedback to help you perceive shapes virtually. Sounds related to your left hand will play in your left ear. Sounds related to your right hand will play in your right ear. One kind of sound gets louder the closer your fingertips are to the shape. A different sound plays when your fingertips are touching or inside the shape. The touching sound gets quieter the closer your fingertips are to the center of the shape. If your fingertips are within 2 cm away from an edge or corner, it will play special edge and corner sounds. Also, the edge and corner sounds get louder the closer your fingertips to edges or corners. For example, the cube has eight corners and 12 edges (Play cube now). The sounds for edge and corner are different (lead the user to hear different sounds). "
 2. Strategies:
 - a. Use both hands to explore the boundary of the shape which is aligned to three axes.
 - b. Use only one fingertip to perceive the axis and learn to maintain vertical to the axis.
- ii. Experiment 1 (distinguishing shapes, 30 minutes). Explain that they will interact with three virtual shapes one at a time in random order. Inform them they will have up to 4 minutes to interact with each shape. Describe the shapes before the experiment begins. Afterwards, ask them which order they interacted with the shapes. Explain that we will repeat with multiple sets of shapes. Take a 3-minute break after each set. The sets are:
 1. Set 1: bottle, bowl, mug

- a. Description: “The bottle is a wine bottle, thin neck; the bowl is a wide opened bowl and the mug has a handle. They are all scaled and fitted inside a box whose size is 0.4 meters.”
- 2. Set 2: sofa, bed, table (scaled to 1/10 size)
 - a. Description: “The sofa is a two-seater sofa with two armrest and back; the bed is a twin size bed with headboard; the table is square shaped and has four legs. They are all scaled and fitted inside a box whose size is 0.4 meters.”
- iii. Experiment 2 (Shape Percept, 16 minutes). Explain that we will give the user 5 minutes to explore a shape and then ask them to touch certain parts of the shape. Tell them that it’s a pyramid with a point on top, four sloping sides, and a bottom. After 5 minutes, ask the user to perform the following tasks (up to 1 minute each):
 1. Touch the top point.
 2. Touch the bottom face.
 3. Touch any bottom corner.
- iv. (10 minutes) Ask the user for their opinion about the software.

B FORMAL STUDY PROTOCOL

- I. Review the Informed Consent Form.
 - i. Review the Health and Safety Warnings for Oculus Quest VR device. Individuals at risk per the Health and Safety Warnings should not participate.
 - ii. Obtain verbal consent.
- II. Explain: “We have created a program for touching virtual 3D shapes that uses only sonification or sound feedback to let you know if your hands are near or touching the shape. Our goal is to evaluate whether our program can allow you to distinguish virtual 3D shapes and how accurately you can perceive the shape. This study will take approximately 1 hour. In the study, you will use our program to touch various objects.”
- III. Direct the participant to wear the VR headset and run the program.
- IV. Procedure:
 - i. Training (10 minutes). Have the user interact with a prism, torus (I choose the mug because it contains all sounds and it’s complicated). Tell them they will have up to 5 minutes to interact with the shape.
 1. Explain the kinds of sonification: “Our project provides four types of sound feedback to help you perceive shapes virtually. Sounds related to your left hand will play in your left ear. Sounds related to your right hand will play in your right ear. One kind of sound gets louder the closer your fingertips are to the shape. A different sound plays when each of your fingertips are touching or inside the shape. (The pitch for each fingertip gets higher from thumb to little finger). The touching sound gets quieter the closer your fingertips are to the center of the shape. If your fingertips are within 2 cm away from an edge or corner, it will play special edge and corner sounds. Also, the edge and corner sounds get louder the closer your fingertips to edges or corners.”
 2. Strategies:
 - a. Use both hands to explore the boundary of the shape which is aligned to three axes.
 - b. Use only one fingertip perceive the axis and learn to maintain be vertical to the axis.
 - c. Move along the axis to perceive the changes in the shape of objects, such as surface protrusions or whether the interior of the object is hollow.
 - ii. Experiment 1 (distinguishing shapes, 15+3 minutes). Explain that they will interact with three virtual shapes one at a time. Inform them they will have up to 5 minutes to interact with each shape. Describe the shapes before the experiment begins. After interacting with each shape, provide them with 3 choices. Ask them to choose which object they interacted with.
 1. Cube (cube, sphere, cone)
 - a. Description: “The shapes are all larger-than-life. They are uniformly scaled to fit inside a cube area whose size is 0.4 meters or 15.75 inches.”
 - b. Visual description, scale, orientation, translate.
 2. Bowl (wine bottle, bowl, mug)
 - a. Description: “The bottle is a wine bottle with a thin neck and no hollow interior. The bowl is a round soup bowl. The mug is a typical coffee mug with a smooth, rounded handle.
 - b. The shapes are all larger-than-life. They are uniformly scaled to fit inside a cube area whose size is 0.4 meters or 15.75 inches.”
 3. Bed (sofa, bed, table)
 - a. Description: “The sofa is a two-seater sofa with a back, armrests on either side, and short feet. The bed is a twin-sized bed with a headboard. The mattress is on a platform with no feet. The table is square shaped and has four legs.
 - b. The shapes are all smaller-than-life. They are uniformly scaled to fit inside a cube area whose size is 0.4 meters or 15.75 inches.”
 - iii. Experiment 2 (Shape Percept, 20 minutes)
 1. Ask the user to touch the following parts without sound feedback (up to 1 minute each):
 - a. A pyramid with a point on top, four sloping sides, and a flat bottom. The pyramid is placed 0.4 meters or 15.75 inches in front of you/oculus. Its base is 1 meter or 3 feet and 3 inches above the ground. The front edge of the pyramid is facing you.
 1. Top point
 2. Bottom face
 3. Any bottom corner
 - b. Cat. The cat is placed 0.4 meters or 15.75 inches in front of you. It is standing upright. Its tail is pointed straight back. It is 1 meter or 3 feet and 3 inches above the ground. It is 0.4 meters or 16 inches tall. The cat is facing to the right.
 1. Either ear
 2. Any paw
 3. Any part of the tail
 2. Give the user five minutes to explore a shape with sound feedback. Ask the user to touch the same parts as above (up to 1 minute each).
 - iv. d. (10 minutes) Ask the user the post-study questions.

Table 1: Sound volume parameters referenced in Section 3. Valid Distance refers to the range in which that sound is played.

Sound	a	b	D (cm)	Valid Distance (cm)
Guidance	0	1	50	$0 < x < 50$
Contact	0	0.8	$ \min(\text{SDF}) $	$0 < x < \min(\text{SDF}) $
Edge/Corner	0.6	0.4	2	$0 < x < 2$

Table 2: Results of a pilot study on landmark localization. Participants were asked to locate three landmarks each on a pyramid and dog. The closest distance from their fingertips to the landmarks is shown in centimeters. Units are in centimeters, and those above 10cm are in blue.

	pyramid			dog		
	top corner	bottom face	bottom corner	nose	feet	Tail
P1*	32.1	1.7	18.0	12.5	0.1	1.8
P2*	4.2	0.1	13.9	7.3	13.4	1.7
P3*	1.9	0.1	1.4	7.3	0.2	0.5
P4*	7.0	0.4	7.6	1.7	2.2	0.0
P5*	0.8	0.1	7.6	6.6	1.6	13.7
P6*	10.7	4.4	1.5	2.8	2.6	6.4

Table 3: Results of a formal study on shape recognition. Users were asked to identify the prompt object from a set of three possibilities, divided into three categories (red, yellow, green columns). Correct choices are shown as ", incorrect in blue. Time to decision is shown in minutes and seconds. Participants chose correctly in 22 out of 27 trials (81%). Participants P1–9 were sighted and P10–15 were BVI. No significant difference was found between sighted and BVI participants.

	m:s	Prompt	Choice	m:s	Prompt	Choice	m:s	Prompt	Choice
P1	2:18	cube	"	4:49	bottle	mug	4:49	table	"
P2	2:18	sphere	"	3:17	bowl	mug	4:53	sofa	"
P3	2:38	cone	"	3:16	mug	"	2:16	bed	"
P4	0:47	cube	"	0:49	bottle	"	1:23	table	"
P5	0:19	sphere	"	1:47	bowl	"	3:41	sofa	"
P6	1:12	cone	"	2:40	mug	"	3:35	bed	sofa
P7	0:20	cube	"	2:00	bottle	"	0:58	table	"
P8	0:20	sphere	"	0:20	bowl	"	1:18	sofa	bed
P9	1:43	cone	"	2:27	mug	bowl	5:10	bed	"
P10	2:37	sphere	"	0:10	bowl	"	0:36	sofa	table
P11	0:28	cube	"	1:50	mug	"	0:34	table	"
P12	0:51	cone	"	1:35	bottle	"	1:24	bed	"
P13	0:11	cube	"	0:05	bottle	"	0:10	table	"
P14	1:35	sphere	"	2:14	bowl	"	3:06	bed	sofa
P15	4:39	cone	cube	0:30	mug	"	4:00	sofa	"

Table 4: Numerical results of our formal study on landmark localization. The decrease in landmark positioning error is shown using ShapeSonic versus no sonification (larger numbers are better). Error is the distance from fingertips to landmark measured in centimeters. The average decrease was 9.6 cm. Participants P1–9 were sighted and P10–15 were BVI. No significant difference was found between sighted and BVI participants. See Figure 6 in the paper for a visualization.

	pyramid			cat		
	top corner	bottom face	bottom corner	ear	feet	tail
P1	3.6	0.4	-1.8	0.2	6.1	1.6
P2	12.0	0.0	3.8	10.0	11.0	18.0
P3	8.8	4.0	8.2	8.8	17.1	8.2
P4	10.1	3.5	13.0	6.5	6.1	2.4
P5	8.2	-1.5	1.5	4.6	6.4	8.3
P6	8.3	10.6	11.9	3.6	6.1	27.3
P7	19.4	2.7	8.1	1.3	-0.3	20.3
P8	20.7	-0.1	10.6	8.2	4.9	11.4
P9	7.9	9.4	3.7	15.3	16.8	24.7
P10	-0.1	1.7	3.5	8.8	2.4	3.4
P11	55.5	10.3	12.9	17.9	23.4	7.0
P12	1.7	12.3	22.6	9.2	8.1	-1.7
P13	14.6	0.6	28.3	20.2	9.5	-5.2
P14	21.6	-4.5	9.2	5.8	0.2	-2.3
P15	2.4	11.0	5.5	31.5	18.4	55.2

Table 5: Results of a pilot study on shape recognition. Participants were asked to identify the order of three presented shapes. Incorrectly ordered objects are marked in blue. In the first set (bowl, mug, and bottle), 2 of 6 participants were entirely correct. In the second set (table, bed, and sofa), 3 of 6 participants were entirely correct. Guessing randomly would have resulted in 1 of 6 entirely correct.

	m:s (average)	Prompt	Choice (#correct shape)	m:s (average)	Prompt	Choice (#correct shape)
P1*	4:24	bowl, mug, bottle	correct (3)	4:37	table, bed, sofa	sofa, bed, table (1)
P2*	3:11	bottle, mug, bowl	mug, bottle, bowl (1)	3:44	table, bed, sofa	correct (3)
P3*	3:37	bowl, bottle, mug	correct (3)	4:06	table, bed, sofa	bed, sofa, table (0)
P4*	3:00	bottle, mug, bowl	mug, bottle, bowl (1)	1:17	sofa, table, bed	correct (3)
P5*	4:16	mug, bowl, bottle	mug, bottle, bowl (1)	4:06	bed, table, sofa	correct (3)
P6*	2:41	bottle, bowl, mug	bowl, bottle, mug (1)	4:08	table, sofa, bed	table, bed, sofa (1)